9.5 An Example: Smoking and Lung Cancer

The fascinating history of the debates over smoking and lung cancer illustrates the difficulties of causal inference and prediction from policy studies, and also illustrates some common mistakes. Perhaps no other hypothetical cause and effect relationship has been so thoroughly studied by non-experimental methods or has so neatly divided the professions of medicine and statistics into opposing camps. The theoretical results of this and the preceding chapters provide some insight into the logic and fallacies of the dispute.

The thumbnail sketch is as follows: In the 1950s a retrospective study by Doll and Hill (1952) found a strong correlation between cigarette smoking and lung cancer. That initial research prompted a number of other studies, both retrospective and prospective, in the United States, the United Kingdom, and soon after in other nations, all of which found strong correlations between cigarette smoking and lung cancer, and more generally between cigarette smoking and cancer and between cigarette smoking and mortality. The correlations prompted health activists and some of the medical press to conclude that cigarette smoking causes death, cancer, and most particularly, lung cancer. Sir Ronald Fisher took very strong exception to the inference, preferring a theory in which smoking behavior and lung cancer are causally connected only through genetics. Fisher wrote letters, essays, and eventually a book against the inference from the statistical dependencies to the causal conclusion. Neyman ventured a criticism of the evidence from retrospective studies. The heavyweights of the statistical profession were thus allied against the methods of the medical community. A review of the evidence containing a response to Fisher and Neyman was published in 1959 by Cornfield, Haenszel, Hammond, Lilienfeld, Shimkin, and Wynder. The Cornfield paper became part of the blueprint for the Report of the Surgeon General on Smoking and Health in 1964, which effectively established that as a political fact smoking would be treated as an unconfounded cause of lung cancer, and set in motion a public health campaign that is with us still. Brownlee (1965) reviewed the 1964 report in the *Journal of the American Statistical Association* and rejected its arguments as statistically unsound for many of the reasons one can imagine Fisher would have given. In 1979, the Surgeon General published a second report on smoking and health, repeating the arguments of the first report but with more extensive data, but offering no serious response to Brownlee's criticisms. The report made strong claims from the evidence, in particular that cigarette smoking was the largest preventable cause of death in the United States. The foreword to the report, by Joseph Califano, was downright vicious, and claimed that any criticism of the conclusions of the report was an attack on science itself. That did not stop P. Burch (1983), a physicist turned theoretical biologist turned statistician, from publishing a lengthy criticism of the second report, again on grounds that were detailed extensions of Fisher's criticisms, but buttressed as well by the first reports of randomized clinical trials of the effects of smoking intervention, all of which were either null or actually suggested

that intervention programs increased mortality. Burch's remarks brought a reply by A. Lilienfeld (1983), which began and ended with an *ad hominem* attack on Burch.

Fisher's criticisms were directed against the claim that uncontrolled observations of a correlation between smoking and cancer, no matter whether retrospective or prospective, provided evidence that smoking causes lung cancer, as against the alternative hypothesis that there are one or more common causes of smoking and lung cancer. His strong views can be understood in the light of features of his career. Fisher had been largely responsible for the introduction of randomized experimental designs, one of the very points of which was to obtain statistical dependencies between a hypothetical cause and effect that could not be explained by the action of unmeasured common causes. Another point of randomization was to insure a well-defined distribution for tests of hypotheses, something Fisher may have doubted was available in observational studies. Throughout his adult life Fisher's research interests had been in heredity, and he had been a strong advocate of the eugenics movement. He was therefore disposed to believe in genetic causes of very detailed features of human behavior and disease. Fisher thought a likely explanation of the correlation of lung cancer and smoking was that a substantial fraction of the population had a genetic predisposition both to smoke and to get lung cancer.

One of Fisher's (1959, p. 8) fundamental criticisms of these epidemiological arguments was that correlation underdetermines causation: besides smoking causing cancer, wrote Fisher "there are two classes of alternative theories which any statistical association, observed without the precautions of a definite experiment, always allows—namely, (1) that the supposed effect is really the cause, or in this case that incipient cancer, or a precancerous condition with chronic inflammation, is a factor in inducing the smoking of cigarettes, or (2) that cigarette smoking and lung cancer, though not mutually causative, are both influenced by a common cause, in this case the individual genotype." Not even Fisher took (1) seriously. To these must be added others Fisher did not mention, for example that smoking and lung cancer have several distinct unmeasured common causes, or that while smoking causes cancer, something unmeasured also causes both smoking and cancer.

If we interpret "statistical association" as statistical dependence, Fisher is correct that given observation only of a statistical dependence between smoking and lung cancer in an uncontrolled study, the possibility that smoking does not cause lung cancer cannot be ruled out. However, he does not mention the possibility that this hypothesis, if true, could have been established without experimentation by finding a factor associated with smoking but independent, or conditionally independent (on variables other than smoking) of cancer. By the 1960s a number of personal and social factors associated with smoking had been identified, and several causes of lung cancer (principally associated with occupational hazards and radiation) potentially independent of smoking had been identified, but their potential bearing on questions of common causes of smoking and lung cancer seems to have gone unnoticed. The more difficult cases to distinguish are the hypotheses that smoking is an unconfounded cause of lung cancer versus the joint

hypotheses that smoking causes cancer and that there is also an unmeasured common cause—or causes—of smoking and cancer.

Fisher's hypothesis that genotype causes both smoking behavior and cancer was speculative, but it wasn't a will-o-the-wisp. Fisher obtained evidence that the smoking behavior of monozygotic twins was more alike than the smoking behavior of dizygotic twins. As his critics pointed out, the fact could be explained on the supposition that monozygotic twins are more encouraged by everyone about them to do things alike than are dizygotic twins, but Fisher was surely correct that it could also be explained by a genetic disposition to smoke. On the other side, Fisher could refer to evidence that some forms of cancer have genetic causes.

The paper by Cornfield et al. (including Lilienfeld) argued that while lung cancer may well have other causes besides, cigarette smoking causes lung cancer. This view had already been announced by official study groups in the United States and Great Britain. Cornfield's paper is of more scientific interest than the Surgeon General's report five years later, in part because the former is not primarily a political document. Cornfield et al. claimed the existing data showed several things:

1. Carcinomas of the lung found at autopsy had systematically increased since 1900, although different studies gave different rates of increase. Lung cancers are found to increase monotonically with the amount of cigarette smoking and to be higher in current than in former cigarette smokers. In large prospective studies diagnoses of lung cancer may have an unknown error rate, but the total death rate also increases monotonically with cigarette smoking.

2. Lung cancer mortality rates are higher in urban than in rural populations, and rural people smoke less than city people, but in both populations smokers have higher death rates from lung cancer than do nonsmokers.

3. Men have much higher death rates from lung cancer than women, especially among persons over 55, but women smoked much less and as a class had taken up the habit much later than men.

4. There are a host of causes of lung cancer, including a variety of industrial pollutants and unknown circumstances associated with socioeconomic class, with the poorer and less well off more likely than the better off to contract the disease, but no more likely to smoke. Cornfield et al. emphasize that "The population exposed to established industrial carcinogens is small, and these agents cannot account for the increasing lung-cancer risk in the remainder of the population. Also, the effects associated with socioeconomic class and related characteristics are smaller than those noted for smoking history, and the smoking class differences cannot be accounted for in terms of these other effects" (p. 179). This passage states that the difference in cancer rates for smokers and nonsmokers

could not be explained by socioeconomic differences. While this claim was very likely true, no analysis was given in support of it, and the central question of whether smoking and lung cancer were independent or nearly independent conditional on all subsets of the known risk factors *that are not effects* of smoking and cancer—area of residence, exposure to known carcinogens, socioeconomic class, and so on, was not considered. Instead, Cornfield et al. note that different studies measured different variables and "The important fact is that in all studies when other variables are held constant, cigarette smoking retains its high association with lung cancer."

5. Cigarette smoking is not associated with increased cancer of the upper respiratory tract, the mouth tissues or the fingers. Carcinoma of the trachea, for example, is a rarity. But, Cornfield et al. point out, "There is no a priori reason why a carcinogen that produces bronchogenic cancer in man should also produce neoplastic changes in the anspharynx or in other sites" (p. 186).

6. Experimental evidence shows that cigarette smoke inhibits the action of the cilia in cows, rats and rabbits. Inhibition of the cilia interferes with the removal of foreign material from the surface of the bronchia. Damage to ciliated cells is more frequent in smokers than in nonsmokers.

7. Application of cigarette tar directly to the bronchia of dogs produced changes in the cells, and in some but not all other experiments applications of tobacco tar to the skin of mice produced cancers. Exposure of mice to cigarette smoke for up to 200 days produced cell changes but no cancers.

8. A number of aromatic polycyclic compounds have been isolated in tobacco smoke, and one of them, the α form of benzopyrene, was known to be a carcinogen.

Perhaps the most original technical part of the argument was a kind of sensitivity analysis of the hypothesis that smoking causes lung cancer. Cornfield et al. considered a single hypothetical binary latent variable causing lung cancer and statistically dependent on smoking behavior. They argued such a latent cause would have to be almost perfectly associated with lung cancer and strongly associated with smoking to account for the observed association. The argument neglected, however, the reasonable possibility of multiple common causes of smoking and lung cancer, and had no clear bearing on the hypothesis that the observed association of smoking and lung cancer is due both to a direct influence and to common causes.

In sum, Cornfield et al. thought they could show a mechanism for smoking to cause cancer, and claimed evidence from animal studies, although their position in that regard tended to trip over itself (compare items 5 and 7). They didn't put the statistical case entirely clearly, but their position seems to have been that lung cancer is also caused by a

number of measurable factors that are not plausibly regarded as effects of smoking but which may cause smoking, and that smoking and cancer remain statistically dependent conditional on these factors. Against Fisher they argued as follows:

> The difficulties with the constitutional hypothesis include the following considerations: (a) changes in lung-cancer mortality over the last half century; (b) the carcinogenicity of tobacco tars for experimental animals; (c) the existence of a large effect from pipe and cigar tobacco on cancer of the buccal cavity and larynx but not on cancer of the lung; (d) the reduced lung-cancer mortality among discontinued cigarette smokers. No one of these considerations is perhaps sufficient by itself to counter the constitutional hypothesis, ad hoc modification of which can accommodate each additional piece of evidence. A point is reached, however, when a continuously modified hypothesis becomes difficult to entertain seriously. (p. 191)

Logically, Cornfield et al. visited every part of the map. The evidence was supposed to be inconsistent with a common cause of smoking and lung cancer, but also consistent with it. Objections that a study involved self-selection—as Fisher and company would object to (d)—was counted as an "ad hoc modification" of the common cause hypothesis. The same response was in effect given to the unstated but genuine objections that the time series argument ignored the combined effects of dramatic improvements in diagnosis of lung cancer, a tendency of physicians to bias diagnoses of lung cancer for heavy smokers and to overlook such a diagnosis for light smokers, and the systematic increase in the same period of other factors implicated in lung cancer, such as urbanization. The rhetoric of Cornfield et al. converted reasonable demands for sound study designs into *ad hoc* hypotheses. In fact none of the evidence adduced was inconsistent with the "constitutional hypothesis."

A reading of the Cornfield paper suggests that their real objection to a genetic explanation was that it would require a very close correlation between genotypic differences and differences in smoking behavior and liability to various forms of cancer. Pipe and cigar smokers would have to differ genotypically from cigarette smokers; light cigarette smokers would have to differ genotypically from heavy cigarette smokers; those who quit cigarette smoking would have to differ genotypically from those who did not. Later the Surgeon General would add that Mormons would have to differ genotypically from non-Mormons and Seventh Day Adventists from nonseventh Day Adventists. The physicians simply didn't believe it. Their skepticism was in keeping with the spirit of a time in which genetic explanations of behavioral differences were increasingly regarded as politically and morally incorrect, and the moribund eugenics movement was coming to be viewed in retrospect as an embarrassing bit of racism.

In 1964 the Surgeon General's report reviewed many of the same studies and arguments as had Cornfield, but it added a set of "Epidemiological Criteria for Causality," said to be sufficient for establishing a causal connection and claimed that smoking and cancer met the criteria. The criteria were indefensible, and they did not

promote any good scientific assessment of the case. The criteria were the "consistency" of the association, the "strength" of the association, the "specificity" of the association, the temporal relationship of the association and the "coherence" of the association.

All of these criteria were left quite vague, but no way of making them precise would suffice for reliably discriminating causal from common causal structures. Consistency meant that separate studies should give the "same" results, but in what respects results should be the same was not specified. Different studies of the relative risk of cigarette smoking gave very different multipliers depending on the gender, age and nationality of the subjects. The results of most studies were the same in that they were all positive; they were plainly not nearly the same in the seriousness of the risk. Why stronger associations should be more likely to indicate causes than weaker associations was not made clear by the report. Specificity meant the putative cause, smoking, should be associated almost uniquely with the putative effect, lung cancer. Cornfield et al. had rejected this requirement on causes for good reason, and it was palpably violated in the smoking data presented by the Surgeon General's report. "Coherence" in the jargon of the report meant that no other explanation of the data was possible, a criterion the observational data did not meet in this case. The temporal issue concerned the correlation between increase in cigarette smoking and increase in lung cancer, with a lag of many years. Critics pointed out that the time series were confounded with urbanization, diagnostic changes and other factors, and that the very criterion Cornfield et al. had used to avoid the issue of the unreliability of diagnoses, namely total mortality, was, when age-adjusted, uncorrelated with cigarette consumption over the century.

Brownlee (1965) made many of these points in his review of the report in the *Journal of the American Statistical Association.* His contempt for the level of argument in the report was plain, and his conclusion was that Fisher's alternative hypothesis had not been eliminated or even very seriously addressed. In Brownlee's view, the Surgeon General's report had only two arguments against a genetic common cause: (a) the genetic hypothesis would allegedly have to be very complicated to explain the dose/response data, and (b) the rapid historical rise in lung cancer following by about 20 years a rapid historical rise in cigarette smoking. Brownlee did not address (a), but he argued strongly that (b) is poor evidence because of changes in diagnostics, changes in other factors of known and unknown relevance, and because of changes in the survival rate of weak neonates whom, as adults, might be more prone to lung cancer.

One of the more interesting aspects of the review was Brownlee's "very simplified" proposal for a statistical analysis of "E_2 causes E_1" which was that E_1 and E_2 be dependent conditional on every possible vector of values for all other variables of the system. Brownlee realized, of course, that his condition did not separate "E_2 causes E_1" from E_1 causes E_2," but that was not a problem with smoking and cancer. But even ignoring the direction of causation, Brownlee's condition—perhaps suggested to him by the fact that the same principle is used (erroneously) in regression—is quite wrong. It

would be satisfied, for example, if, E_1 and E_2 had no causal connection whatsoever provided some measured variable E_j were a direct effect of both E_1 and E_2.

Brownlee thought his way of considering the matter was important for prediction and intervention:

> If the inequality holds only for, say, one particular subset E_j,..., E_k, and for all other subsets equality holds, and if the subset E_j,...,E_k occurs in the population with low probability, then $\Pr\{E_1|E_2\}$, while not strictly equal to $\Pr\{E_1|E_2^c\}$, will be numerically close to it, and then E_2 as a cause of E_1 may be of small practical importance. These considerations are related to the Committee's responsibility for assessment of the *magnitude* of the health hazard (page 8). Further complexities arise when we distinguish between cases in which one of the required secondary conditions E_j,...,E_k is, on the one hand, presumably controllable by the individual, e.g., the eating of parsnips, or uncontrollable, e.g., the presence of some genetic property. In the latter case, it further makes a difference whether the genetic property is identifiable or non-identifiable: for example it could be brown eyes which is the significant subsidiary condition E_j, and we could tell everybody with not-brown eyes it was safe for *them* to smoke. (p. 725)

No one seems to have given any better thought than this to the question of how to predict the effects of public policy intervention against smoking. Brownlee regretted that the Surgeon General's report made no explicit attempt to estimate the expected increase in life expectancy from not smoking or from quitting after various histories.

Fifteen years later, in 1979, the second Surgeon General's Report on Smoking and Health was able to report studies that showed a monotonic increase in mortality rates with virtually every feature of smoking practice that increased smoke in the lungs: number of cigarettes smoked per day, number of years of smoking, inhaling versus not inhaling, low tar and nicotine versus high tar and nicotine, length of cigarette habitually left unsmoked. The monotonic increase in mortality rates with cigarette smoking had been shown in England, the continental United States, Hawaii, Japan, Scandinavia and elsewhere, for whites and blacks, for men and women. The report dismissed Fisher's hypothesis in a single paragraph by citing a Scandinavian study (Cederlof, Friberg, and Lundman 1977) that included monozygotic and dizygotic twins:

> When smokers and nonsmokers among the dizygotic pairs were compared, a mortality ratio of 1.45 for males and 1.21 for females was observed. Corresponding mortality ratios for the monozygotic pairs were 1.5 for males and 1.222 for females. Commenting on the constitutional hypothesis and lung cancer, the authors observed that "the constitutional hypothesis as advanced by Fisher and still supported by a few, has here been tested in twin studies. The results from the Swedish monozygotic twin series speak strongly against the constitutional hypothesis."

The second Surgeon General's report claimed that tobacco smoking is responsible for 30% of all cancer deaths; cigarette smoking is responsible for 85% of all lung cancer deaths.

A year before the report appeared, in a paper for the British Statistical Association P. Burch (1978) had used the example of smoking and lung cancer to illustrate the problems of distinguishing causes from common causes without experiment. In 1982 he published a full fledged assault on the second Surgeon General's report. The criticisms of the argument of the report were similar to Brownlee's criticisms of the 1964 report, but Burch was less restrained and his objections more pointed. His first criticism was that while all of the studies showed a increase in risk of mortality with cigarette smoking, the degree of increase varied widely from study to study. In some studies the age adjusted multiple regression of mortality on cigarettes, beer, wine and liquor consumption gave a smaller partial correlation with cigarettes than with beer drinking. Burch gave no explanation of why the regression model should be an even approximately correct account of the causal relations. Burch thought the fact that the apparent dose/response curve for various culturally, geographically, and ethnically distinct groups were very different indicated that the effect of cigarettes was significantly confounded with environmental or genetic causes. He wanted the Surgeon General to produce a unified theory of the causes of lung cancer, with confidence intervals for any relevant parameter estimates: Where, he asked, did the 85% figure come from?

Burch pointed out, correctly, that the cohort of 1487 dizygotic and 572 monozygotic twins in the Scandinavian study born between 1901 and 1925 gave no support at all to the claim that the constitutional explanation of the connection between smoking and lung cancer had been refuted, despite the announcements of the authors of that study. The study showed that of the dizygotes exactly 2 nonsmokers or infrequent smokers had died of lung cancer and 10 heavy smokers had died of lung cancer; of the monozygotes, 2 low non smokers and 2 heavy smokers had died of the disease. The numbers were useless, but if they suggested anything, it was that if genetic variation was controlled there is no difference in lung cancer rates between smokers and nonsmokers. The Surgeon General's report of the conclusion of the Scandinavian study was accurate, but not the less misleading for that.

Burch also gave a novel discussion of the time series data, arguing that it virtually refuted the causal hypothesis. The Surgeon General and others had used the time series in a direct way. In the U.K. for example, male cigarette consumption per capita had increased roughly a hundredfold between 1890 and 1960, with a slight decrease thereafter. The age-standardized male death rate from lung cancer began to increase steeply about 1920, suggesting a thirty-year lag, consistent with the fact that people often begin smoking in their twenties and typically present lung cancer in their fifties. According to Burch's data, the onset of cigarette smoking for women lagged behind males by some years, and did not begin until the 1920s. The Surgeon General's report noted that the death rate from lung cancer for women had also increased dramatically

between 1920 and 1980. Burch pointed out that the autocorrelations for the male series and female series didn't mesh: there was no lag in death rates for the women. Using U.K. data, Burch plotted the *percentage change* in the age-standardized death rate from lung cancer for both men and women from 1900 to 1980. The curves matched perfectly until 1960. Burch's conclusion is that whatever caused the increase in death rates from lung cancer affected both men and women at the same time, from the beginning of the century on, although whatever it is had a smaller absolute effect on women than on men. But then the whatever-it-was could not have been cigarette smoking, since increases in women's consumption of cigarettes lagged twenty to thirty years behind male increases.

Burch was relentless. The Surgeon General's report had cited the low occurrence of lung cancer among Mormons. Burch pointed out that Mormon's in Utah not only have lower age-adjusted incidences of cancer than the general population, but also have higher incidences than non-Mormon nonsmokers in Utah. Evidently their lower lung cancer rates could not be simply attributed to their smoking habits.

Abraham Lilienfeld, who only shortly before had written a textbook on epidemiology and who had been involved with the smoking and cancer issue for more than twenty years, published a reply to Burch that is of some interest. Lilienfeld gives the impression of being at once defensive and disdainful. His defense of the Surgeon General's report began with an *ad hominem* attack, suggesting that Burch was so out of fashion as to be a crank, and ended with another *ad hominem*, demanding that if Burch wanted to criticize others' inferences from their data he go get his own. The most substantive reply Lilienfeld offered is that the detailed correlation of lung cancer with smoking habits in one subpopulation after another makes it seem very implausible that the association is due to a common cause. Lilienfeld said, citing himself, that the conclusion that 85% of lung cancer deaths are due to cigarettes is based on the relative risk for cigarette smokers and the frequency of cigarette smoking in the population, predicting, in effect, that if cigarette smoking ceased the death rate from lung cancer would decline by that percentage. (The prediction would only be correct, Burch pointed out in response, provided cigarette smoking is a completely unconfounded cause of lung cancer.) Lilienfeld challenged the source of Burch's data on female cigarette consumption early in the century, which Burch subsequently admitted were estimates.

Both Burch and Lilienfeld discussed a then recent report by Rose et al. (1982) on a ten-year randomized smoking intervention study. The Rose study, and another that appeared at nearly the same time with virtually the same results, illustrates the hazards of prediction. Middle-aged male smokers were assigned randomly to a treatment or nontreatment group. The treatment group was encouraged to quit smoking and given counseling and support to that end. By self-report, a large proportion of the treatment group either quit or reduced cigarette smoking. The difference in self-reported smoking levels between the treatment and nontreatment groups was thus considerable, although the difference declined toward the end of the ten-year study. To most everyone's dismay, Rose found that there was no statistically significant difference in lung cancer between

the two groups after ten years (or after five), but there was a difference in overall mortality—the group that had been encouraged to quit smoking, and had in part done so, suffered higher mortality.

Fully ignoring their own evidence, the authors of the Rose study concluded nonetheless that smokers should be encouraged to give up smoking, which makes one wonder why they bothered with a randomized trial. Burch found the Rose report unsurprising; Lilienfeld claimed the numbers of lung cancer deaths in the sample are too small to be reliable, although he did not fault the Surgeon General's report for using the Scandinavian data, where the numbers are even smaller, and he simply quoted the conclusion of the report, which seems almost disingenuous. To Burch's evident delight, as Lilienfeld's defense of the Surgeon General appeared so did yet further experimental evidence that intervening in smoker's behavior has no benign effect on lung cancer rates. The Multiple Risk Factor Intervention Trial Research Group (1982) reported the results after six years of a much larger randomized experimental intervention study producing roughly three times the number of lung cancer deaths as in the Rose study. But the intervention group showed more lung cancer deaths than the usual care group! The absolute numbers were small in both studies but there could be no doubt that nothing like the results expected by the epidemiological community had materialized.

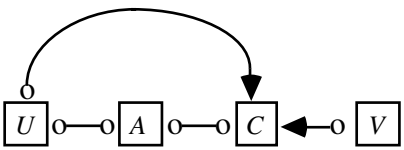
The results of the controlled intervention trials illustrate how naive it is to think that experimentation always produces unambiguous results, or frees one from requirements of prior knowledge. One possible explanation for the null effects of intervention on lung cancer, for example, is that the reduced smoking produced by intervention was concentrated among those whose lungs were already in poor health and who were most likely to get lung cancer in any case. (Rose et al. gave insufficient information for an analysis of the correlation of smoking behavior and lung cancer within the intervention group.) This possibility could have been tested by experiments using blocks more finely selected by health of the subjects.

In retrospect the general lines of the dispute were fairly simple. The statistical community focused on the want of a good scientific argument against a hypothesis given prestige by one of their own; the medical community acted like Bayesians who gave the "constitutional" hypothesis necessary to account for the dose/response data so low a prior that it did not merit serious consideration. Neither side understood what uncontrolled studies could and could not determine about causal relations and the effects of interventions. The statisticians pretended to an understanding of causality and correlation they did not have; the epidemiologists resorted to informal and often irrelevant criteria, appeals to plausibility, and in the worst case to *ad hominem*.

Fisher's prestige as well as his arguments set the line for statisticians, and the line was that uncontrolled observations cannot distinguish among three cases: smoking causes cancer, something causes smoking and cancer, or something causes smoking and cancer and smoking causes cancer. The most likely candidate for the "something" was genotype. Fisher was wrong about the logic of the matter, but the issue never was satisfactorily

clarified, even though some statisticians, notably Brownlee and Burch, tried unsuccessfully to characterize more precisely the connection between probability and causality. While the statisticians didn't get the connection between causality and probability right, the Surgeon General's "epidemiological criteria for causality" were inadequate and arguments in defense of the conclusions of the Surgeon General's Report were flawed. The real view of the medical community seems to have been that it was just too implausible to suppose that genotype strongly influenced how much one smoked, whether one smoked at all, whether one smoked cigarettes as against a cigar or pipe, whether one was a Mormon or a Seventh day Adventist, and whether one quit smoking or not. After Cornfield's survey the medical and public health communities gave the common cause hypothesis more invective than serious consideration. And, finally, in contrast to Burch, who was an outsider and maverick, leading epidemiologists, such as Lilienfeld, seem simply not to have understood that if the relation between smoking and cancer is confounded by one or more common causes, the effects of abolishing smoking cannot be predicted from the "risk ratios," that is, from sample conditional probabilities. The subsequent controlled smoking intervention studies gave evidence of how very bad were the expectations based on uncontrolled observation of the relative risks of lung cancer in those who quit smoking compared to those who did not.

9.6 Appendix



(io*)

Figure 9.18

We will prove that the partially oriented inducing path graph (io*) in figure 9.18, together with the assumptions that U causes A, that there is no common cause of U and C, and that every directed path from U to C contains A, entail that A causes C and that there is a latent common cause of A and C. We assume that A is not a deterministic function of U.

Let $\mathbf{O} = \{A,C,U,V\}$, and G be the directed acyclic graph that generated (io*). The U o$\rightarrow$ C edge in (io*) entails that in the inducing path graph of G either $U \rightarrow C$ or $U \leftrightarrow C$. If there is a $U \leftrightarrow C$ edge, then there is a latent common cause of U and C, contrary to our assumption. Hence the inducing path graph contains a $U \rightarrow C$ edge. It follows that in G there is a directed path from U to C. Because every directed path from U to C contains A, there is a directed path from A to C in G. Hence A causes C.